Getting Your Money's Worth: Diagnostic Benchmarking for Commissioning

Les Lambert, Lambert Engineering

ABSTRACT

Commissioning is important for new or modified facilities to achieve expected efficiency. Inspections and spot tests can ensure equipment has needed capabilities. But demonstrating continuing operational efficiency requires observing operation over time during and after warranty. Valid comparisons of modeled or measured baseline energy use to post-construction or post-modification data are needed to prove that operational efficiency has been achieved.

Diagnostic benchmarks enable such comparisons, providing accurate, economical baselines using whole-building simulated or metered data. They spot performance shortfalls, support speedy problem identification and resolution, and can ensure sustainable efficiency after commissioning.

Diagnostic benchmarks are graphs that show trends of average energy use rates, versus coincident average outside temperature, forming baselines. Post-construction, weekly measurements optimize response time and accuracy. Cyclic integrals of energy use rate, over one week, minimize occupancy-driven rate fluctuations, and preserve weekly time resolution. Trends reveal weather dependency.

Post-construction data above simulated baseline, or showing less-than-expected reductions, indicate performance shortfalls. Data at or below simulated baseline indicates successful performance. If post-construction shortfalls occur, corrections are assessed by whether subsequent weekly data points approach the expected levels.

Examples from a military installation show:

- Five-fold initial difference between simulated LEED design and actual performance.
- Shortcomings in contractual commissioning mechanisms.
- Detection and correction of inappropriate operation.
- Pre-testing of "soft starts" to support fan scheduling.
- Reductions in gas and electric use after correction of over-ventilation attributable to insufficient VAV throttling range and plenum leakage.
- Tuning HVAC controls to minimize reheat and over-ventilation.

Introduction and Background

Measurement-based continuous energy performance metrics are needed to achieve and sustain efficiency in buildings. Any investment or effort to enhance energy efficiency in buildings warrants commissioning, based on such metrics, for five important reasons:

- Ensure that the enhancements are correctly conceived, designed, and implemented.
- Measure performance, to ensure performance levels or expected savings are achieved.
- Get feedback for all concerned, so the next project can be more effective.
- Achieve sustainable savings over the project's life, despite equipment and human failure.
- Assure energy efficiency financiers and regulators they are getting their money's worth.

Current design and commissioning[1] practices may include simulations, project tracking to verify design intent is followed, and final inspections and spot tests – including extensive testing and balancing ("TAB"). These practices help guide design, and to ensure specified equipment is correctly installed and delivers specified performance. But they don't prove the design or hardware is suitable, or properly controlled, operated, and maintained for efficiency over time.

As a remedy, we need reliable, inexpensive, objective measurements to characterize efficiency and efficiency improvements. Such measurements should apply to commissioning new buildings, energy efficiency remodels, retrofits, Operating and Maintenance ("O&M") enhancements, and to sustainability - analogous to regular check-ups used by doctors to keep people healthy. Doctors ***always*** check standard, objective, quick diagnostic benchmarks - weight, pulse, blood pressure, and temperature. Changed or out-of-range values get prompt attention.

Similar objective, measurement-based diagnostic benchmarks have been devised for buildings. These benchmarks show if a new, remodeled, re-tuned, or retrofitted building meets, and continues to meet, efficiency expectations.

This paper describes the benchmarks, and presents example applications from a military base. Examples show new building commissioning and a variety of other relevant situations, covering selected process issues and results. These benchmarks have proven effective for detecting and diagnosing problems, and tracking corrections, during continuous commissioning. They are powerful tools for getting your money's worth during initial commissioning.

The key finding on diagnostic benchmarks is, if occupancy-driven fluctuations in energy use (from schedules, temperature setbacks, holidays, etc.) are analytically minimized, resulting smoothed energy data is usually highly correlated with outside temperature. ***Consistent*** operation (i.e. same schedules, temperature settings, etc. from week to week) produces such conditions. Where a robust correlation between smoothed energy use and outside temperature is observed, a benchmark, and changes – deliberate or otherwise – quickly become apparent and measurable.

The theory that leads to this conclusion is:

For buildings in constant physical condition, and consistently occupied and operated, long-term energy use rate varies solely in response to changes in weather.

Benchmark Uses, Requirements, and Characteristics

Diagnostic benchmarks will be shown to be the most quickly responsive, accurate benchmarks available, making them ideal for continuous commissioning of buildings, cradle to grave. Besides measuring changes, they help to diagnose problems, prioritize buildings for retro-commissioning, ensure sustainable savings, accurately size retrofit equipment, and identify and carry out O&M measures that are seldom recognized, let alone executed correctly.

To be effective for such use, diagnostics must be responsive to problems - and associated energy use patterns - due to many causes. To be practical and all-purpose, the diagnostics must deal directly, and sensitively, with metered whole-building energy use.

But comparisons of whole-building energy use are usually perceived as prone to large measurement errors. The ***perception*** of high "noise levels" due to uncontrollable variables such as weather and occupancy causes evaluators and analysts to impose large error bars. For example, proving a 10% reduction, using before-and-after values each with 20% uncertainty, is challenging. Comparing performance of a completed LEED building to pre-construction simulations used to be highly uncertain, because simulated and actual weather weren't the same.

[1] "Commissioning", in this paper, includes re-commissioning, retro-commissioning and continuous commissioning.

Analytical methods are shown to improve the "signal to noise ratio", smoothing energy use data to overcome this concern. The resulting diagnostic benchmarks are useful in near real time for individual buildings, despite occupancy and weather driven variations in energy use.

Building diagnostic benchmarks use only readily available data – building size, energy meter readings, and nearby ambient temperature. Properly analyzed, these data can minimize confusion from occupancy patterns and weather, to instead show energy use patterns that are inherent to the buildings ***as they are operated***. Especially for consistently operated buildings, these benchmarks show substantially improved statistical explanatory power.

The statistical power of these benchmarks is further improved when energy data are dis-aggregated such that heating and cooling energy use for HVAC are recorded on different meters. Many of the examples show buildings where smoothed aggregated energy use is nearly flat, versus outside temperature, or varies substantially less, compared to dis-aggregated energy use.

Using the benchmarks, it is important to maintain separate trends for different operating conditions. For example, known equipment or control faults, expected to produce "outlier" data points, should be trended separately. Trends are sufficiently sensitive to changed operation that you'll usually know, due to new outlier data points, when something has changed from one week to the next. Good data trends tend to be self-evident, and self-validating, by showing data for weeks with atypical operation due to holidays to be outliers. Don't discard atypical data; that's heresy to a statistician. But trend it separately. You can learn important things from outliers.

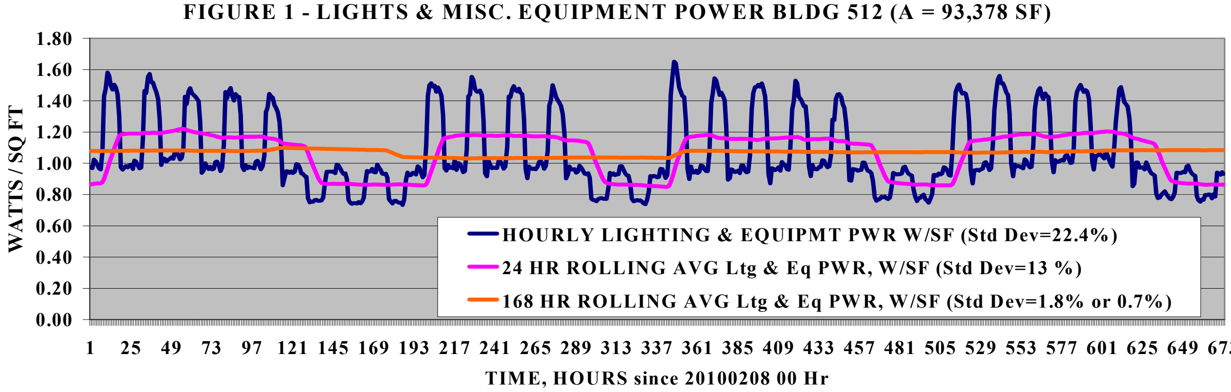
Because benchmarks rapidly measure patterns and effects of operational changes, they serve as effective precursors to energy audits. Walk-through audits often ignore or overlook possible O&M measures. With prior benchmarking, auditors can focus on the most relevant occupancy and control patterns. Benchmarks also enable persistence of savings, providing alarm signals to constrain performance degradation due to equipment failure and human error.

Finally, since diagnostic benchmarks rely on whole-building energy use, and are inherently weather normalized, they account for interactive effects of treatments. Treatment cumulative effects on heating and cooling energy consumption, which are difficult to predict, are revealed. They measure energy effects of changes that are difficult to measure directly – such as envelope improvements, changes to ventilation, and minimizing reheat. As will be shown, measurements previously viewed as impossible or very challenging become more practical.

Reducing Noise from Cyclic Occupancy

Buildings show occupancy-driven cyclic energy use. There are obvious diurnal cycles, but man-made weekly cycles are even more important. Figure 1 shows hourly measured electrical use for lights and miscellaneous equipment in an office facility, over four weeks.

FIGURE 1 - LIGHTS & MISC. EQUIPMENT POWER BLDG 512 (A = 93,378 SF)



deviation, +/- 22.4 percent, is an artifact of using hourly data; it defies our intuition of consistent use. Nearly all that variation, and much of the variation in other uses, is due to cyclic operation.

Finding any randomly chosen hour, day, or month that precisely represents the building's average energy use rate is subject to large sampling error. Avoiding such sampling error requires averaging over exactly one - or an integer number – of cycles. We must divide the cyclic integral of energy use by the cycle duration, to get a representative average energy use rate.

Plotting the rolling average ("RA") energy use rate on a daily and weekly basis in Figure 1 (pink and orange lines) shows how variability is reduced by proper sampling. Using daily (24 hour RA) and weekly (168 hour RA) data reduces the variability to 13% (daily) or to 1.8% (weekly). If the second week (the atypical one, following the three day weekend) is excluded from the weekly averages, the average for the three remaining weeks is 1.076, and the variability is 0.7%, about 1/32nd of the perceived variability of hourly data. This latter weekly average is also more accurate, and useful, than values from monthly utility bills, for two reasons.

First, utility bills are for 30 or 31 days, and may include holidays. In figure 1, weekends and holidays had lower energy use. Adding a weekend with a holiday to this data series would skew the monthly "average" to the low side (1.05). Conversely, adding three weekdays would skew the "average" to the high side (1.08). An arbitrarily selected month might show a low or high value, but ***never*** a value without sampling error. For this example, the average of monthly values would range from 1.05 to 1.08, producing a variability about 4 times greater than the standard deviation for "consistent" weeks (0.7%), those without holidays.

Second, correlation with temperatures works best with more data points. Twenty weeks provide robust baselines in example 3. Sampling error aside, monthly bills need 20 months to achieve the same statistical power. Changed energy use (equipment failures, changed occupancy or operation) that mask changes of interest are more likely in 20 months than in 20 weeks.

Accounting for Temperature

It has been recognized that energy use and weather are correlated since the invention of degree-days. More recently, empirical correlations of energy use and temperature have been used to characterize buildings using monthly utility bills, for example (Goldberg 1980). But the typical utility bill spans a non-integer number of weekly cycles. They do not support weekly sampling, let alone segregation of energy use during holiday or other atypical weeks.

The author presented a benchmarking method, derived from energy balance considerations rather than empirical statistical analysis, that trended cycle-average energy use rate versus coincident temperature difference (Lambert 1998). This benchmarking method originally required both outdoor and indoor temperatures. The author has since determined that for consistently operated buildings, outdoor temperature alone may be used in lieu of indoor-outdoor delta T. This simplification is convenient and produces useful results, as shown here.

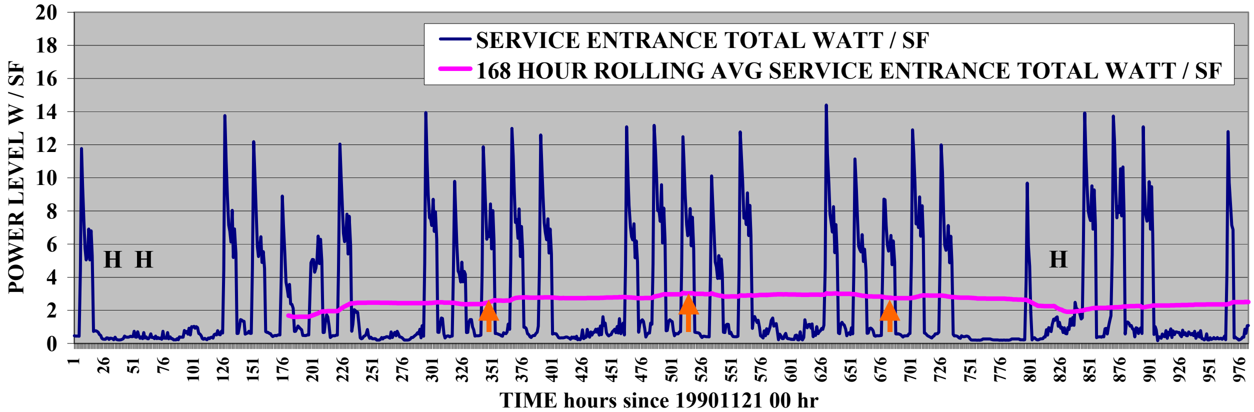
Other benchmarking tools, such as the U. S. Environmental Protection Agency's Energy Star Portfolio Manager ("ESPM") also perform weather normalization for benchmarking of building energy use (EPA 2011). ESPM is becoming the de-facto choice for energy benchmarking for buildings in the US. ESPM, for user convenience, also uses utility billing information, and a small subset (about 157 locations within the US) of the readily available weather data. But choices driven by user convenience constrain ESPM's usefulness.

Synchronizing utility bills with monthly weather files forces approximations, allocating average daily use to specific months. The resulting annual segmented linear fits of billing information to semi-local weather are based on roughly four to six points each for cooling and for heating. These fits show energy - temperature correlations on the order of R^2~0.80, leaving much more unexplained variation in energy use than results from using "typical" weekly data. Weather normalization in ESPM happens in the background, so step changes in trend on a graph are not visible. Therefore, up to a year is needed to see the full impact of a change, using ESPM.

Since ESPM doesn't suit our needs, we'd like to examine how energy use rate changes as outside temperature varies, ***with all else held constant.*** The desired result should be a graph, the slope of which shows how energy use rate varies solely with outside temperature, or $\partial\dot{\mathbf{E}}/\partial\mathbf{T}_{\mathbf{out}}$.

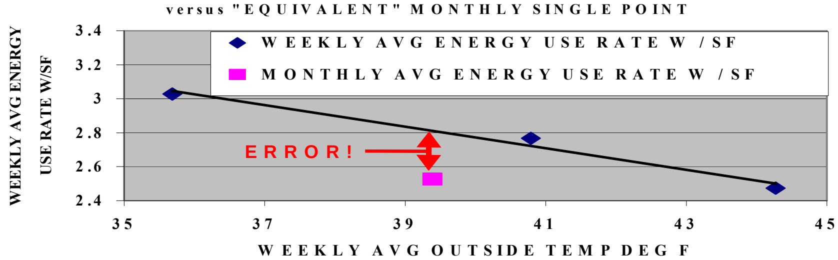
Consider a few weeks of hourly whole-building data, to see the combined effects of occupancy and weather. Figure 2 shows whole-building energy use for a large (199,489 ft^2) all-electric school building in Massachusetts. Note that the data varies over a large range, due mostly to HVAC energy use, which is near zero during nights, weekends, and holidays (H). Also, the data interval involves two major holidays. We must recognize that weeks with holidays or other inconsistent operating patterns do not satisfy "***with all else held constant".***

FIGURE 2 - PCIS TOTAL POWER



The "energy week" used here is from Wednesday noon to Wednesday noon. The example has three weeks of data for "consistent" operation, relatively unaffected by holidays. These values (arrows in figure 2), with coincident outside temperatures, are plotted in figure 3.

FIGURE 3 - WEEKLY TRENDED DATA
versus "EQUIVALENT" MONTHLY SINGLE POINT



Had we characterized December as a monthly average, we would have one monthly value (2.52 Watt/SF), lowered by effects of a holiday. But weekly data, less affected by inconsistent conditions, gives us three data points that show a trend, from one month's data. The "monthly" point for December is in error, 10% below the trendline. If we were using monthly data to get an annualized curve, this monthly point (and a similar point for January, probably including two holidays) would, by virtue of being the coldest months, become "influential outliers" distorting the annual energy use versus outside temperature plot.

Estimating energy use changes. Now we can develop baselines for individual buildings or a complex. If we collect consistent weekly data over a wide-enough outside temperature range, and trend it, we can project energy use for an average weather year at "consistent conditions":

$$E_{\mathbf{annual}} = \text{area} * (\sum \dot{E}_{\mathbf{i}} (T_{\mathbf{i}}) * \text{Hours}_{\mathbf{i}})$$

Summation is over 12 months, with hourly energy use rate, $\dot{E}_{\mathbf{i}}$, determined by pick-off from a curve fit of weekly energy use versus $\mathbf{T_{out}}$, at monthly average temperature $T_{\mathbf{i}}$.

For impact measurement, post-treatment data over a wide range of ambient temperatures should be collected. The post-treatment trend gives a new value for $E_{\mathbf{annual}}$ to compare with baseline. Expected inconsistencies from holidays, etc., can be adjusted for, if needed.

About multiple fuels. Different fuels for heating and cooling allow seeing dis-aggregated effects. Plot and trend them separately, but on the same graph, despite "clutter", so interactive effects can be anticipated or observed. Plotting heating fuel separately made problems such as reheat or failure to turn off heat in the summer easier to spot. In all-electric facilities, if separate metering of heat can be done cheaply, it will likely pay for itself in increased information yields.

Consistent units must be used, so both electrical and gas energy are shown in BTU/hr-ft^2 (sometimes, "BTUh/SF"). The following graphs show gas as if it provides useful heat at 80% efficiency. That's arbitrary. Some buildings had multiple boilers, with nameplate efficiencies ranging from 80%, to "up to 100%". Useful heat estimates were desirable since energy used for reheat can appear as cooling load. But striving for exact efficiencies can be frustrating.

Energy use reflects ***system efficiency***, meter to load. A high-efficiency boiler driving under-slab radiant heat could have a system efficiency of 50%, due to ground losses. A make-up air heater, delivering 100% outside air, used for heating, has low "efficiency" in cold weather.

About Data Collection. Previous graphs use hourly data, but hourly isn't essential. Data in the examples were collected with "eyeball" meter readings, read on Wednesdays; our goal was to read meters once every 168 hours, +/- one hour. Nearby weather is generally available hourly.

With automated metering, ask for weekly reads, Wednesday noon to Wednesday noon.

About the Example Applications

The military mission was saving energy quickly - not flawless measurements. We did continuous benchmarking, for mid-course guidance on buildings that typically had not been thoroughly commissioned under warranty, and therefore needed retro-commissioning. Low electric and gas rates forced low-cost, fast payback O&M upgrades to become a priority.

O&M changes, once identified and agreed to, were put into practice. We attempted to evaluate one change at a time, but there wasn't always time for a full baseline, or a full post-treatment trend. Savings estimates for O&M measures were at least measurement-based, rather than assumption-based. Frequent unexpected outcomes showed measurement-based estimates to be a large improvement over estimates based solely on assumptions and engineering principles.

Also, retro-commissioning to adapt buildings to new missions was sometimes needed. So we were seeking a wide range of problems; design errors, construction errors and omissions, equipment failures, need for facility upgrades, non-optimum building automation system (BAS) use, and human errors. Older buildings obviously needed upgrades like insulation and new boilers. These were the easy answers. We spent more effort on newer buildings, and on observing how old buildings (50% with BAS) were operated, to spot enhanced O&M strategies.

While optimizing O&M, physical plant shortcomings that precluded efficient O&M became apparent. Spending to remedy such weak links was then considered, with a better knowledge of retrofit equipment sizing needs, and how it would benefit operational efficiency. And we needed to detect and fix energy waste due to malfunctions and human errors, quickly.

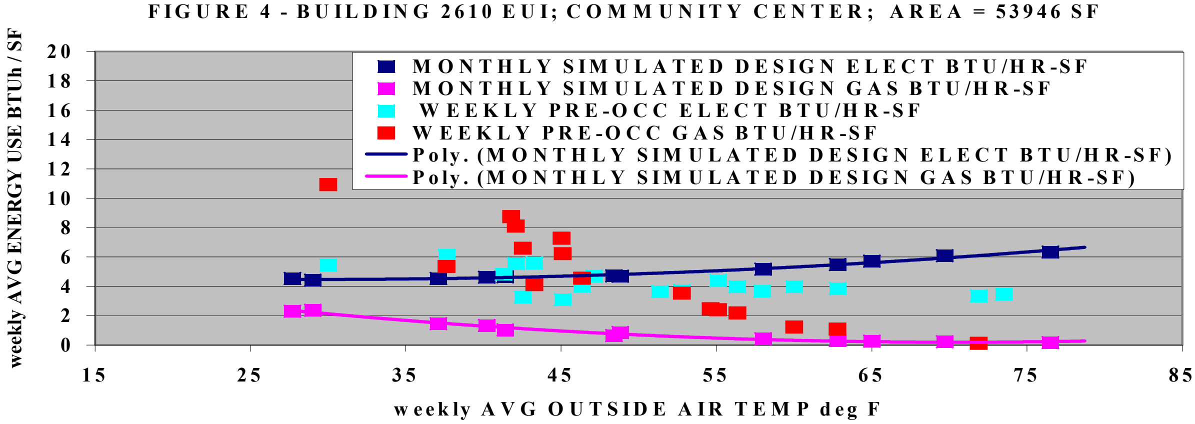
The examples presented aren't intended to show precise annualized measurements. The point is, they demonstrate objective ways to improve measurement feedback, accuracy and speed, which is important to commissioning. Most examples show a statistically robust baseline, and measurement of changes that are analytically challenging, especially with interactive effects.

Example 1: Commissioning a New Building

Building 2610 had suffered a roof failure. An essentially new building was designed, re-using the old slab and walls. Military policy mandated certifiability to LEED silver standards. Modeling to show LEED compliance was required, and the contract required "commissioning". But we already were working on many recent buildings with problems resulting from imperfect initial commissioning. At this point, building design was completed, and a construction contract had been awarded; commissioning provisions consisted of extended TAB, performed before turn-over of the building to the military. We set out to learn how to improve the process.

Figure 4 shows results of the final LEED-required performance simulations for the building, as monthly energy use rates trended with coincident monthly temperatures. Newly commissioned meters were used for weekly readings starting March 2011, also in Figure 4.

FIGURE 4 - BUILDING 2610 EUI; COMMUNITY CENTER; AREA = 53946 SF



The extended TAB was successfully completed. But the early gas data showed gas use rates five-fold above simulations. Delays installing Government Furnished Equipment delayed turnover until July 1, with move-in starting in mid-July. Before July, the military did not "own" the building, and had no contractual standing to request correction of the excessive energy use under warranty. The contractor said the controls bid covered programming initial EMCS schedules ***once***, after full occupancy, during maximum cooling conditions, and then re-tuning once, during heating conditions. The base HVAC and EMCS shops were under a standard injunction from base contracting – first year, tinkering voids the warranty. The shops weren't trained on the building's automation system (a new type on the base) yet anyway.

Later, Contracting decided that occupancy had started 4/15/2011. It appeared full occupancy was likely in late fall (after maximum cooling), and full EMCS scheduling and tuning might not occur until summer 2012 - now deemed to be post-warranty. Things didn't look good:

- Our use of meter data successfully exposed serious performance shortfalls, BUT,
- Contracting mechanisms to secure corrections under warranty weren't in place.
- Corrections weren't going to happen in time to verify with meter data during warranty.
- Proper scheduling *might* correct the high gas use – meanwhile, high gas use may have masked other problems needing to be discovered in time for corrections under warranty.
- The building was about to become a "retro-commissioning opportunity" – among many.
- The base shops had no budgets, methods, or staff for retro-commissioning.

These problems occurred despite successful TAB completion. But work at the base had shown that proper O&M can be more influential in achieving energy-efficient performance than having the right hardware. Diagnostic benchmarking can help point the way to, and verify, efficient O&M. But it must be used in concert with contracting mechanisms that require early initial BAS programming, and support "on-the-fly" adjustments before the end of the warranty.

The exact contracting mechanism needed is beyond the scope of this paper. However, the contract should require demonstration of one year of efficient operation (as shown by diagnostic benchmarks), subsequent to completion of initial scheduling and balancing. This would have corrected most of the problems with the above situation. The initial scheduling and balancing should use the same occupancies and schedules used in pre-construction simulations.

If the building users later request different occupancies or schedules, simulations should be re-run accordingly. This would have removed the time lag awaiting full occupancy during worst case cooling or heating conditions, and motivated cooperation to speed completion, instead of encouraging bid strategies to minimized up-front costs. It also helps to avoid controversy over whether simulated and actual occupancy differ.

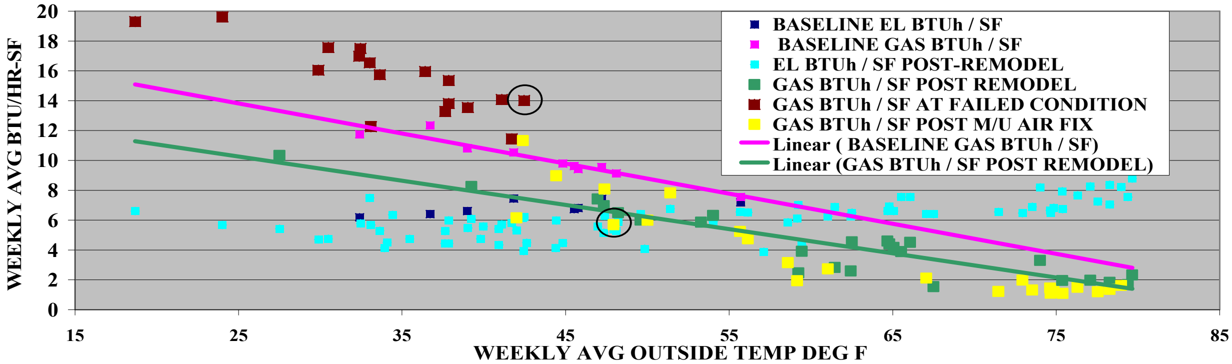
Suitable contracting mechanisms should be achievable. Success showing efficient operation then depends on diagnostic benchmarks working as claimed. The rest of this paper shows how diagnostic benchmarks quickly make valid measurements of efficiency and changes in efficiency for remodels, tune-ups, and for retro-commissioning of already-occupied buildings.

Remaining examples had no LEED simulations to use as performance targets. We used "self benchmarking" and peer comparisons as starting points. For O&M upgrades and "Sustainment, Restoration and Modernization" (SRM) remodels, we worked to accomplish post-modification energy use lower than shown pre-modification by the same building. After optimizing O&M, we recommended investments to remedy physical plant shortcomings, and continued benchmarking to forestall backsliding due to equipment failures and human error.

Example 2: Commissioning an HVAC Remodel

Building 195, the Officers' and NCO Club, was built in 1977. Its aging HVAC system was a maintenance headache. The remodel was intended as an SRM, but also turned out to improve efficiency. An old high-velocity duct system, using a main air handler with mechanical inlet vane VAV, was replaced with lower velocity ducting and variable frequency drives. Boilers and parts of the kitchen make-up air system were also upgraded. The upgraded BAS was extended to control and monitor new equipment, kitchen ventilation excepted. We had time to collect a partial baseline before the remodel started, see Figure 5.

FIGURE 5 - WEEKLY EUI - BLDG 195 NCO/OFFICER CLUB, AREA = 39,317 SF



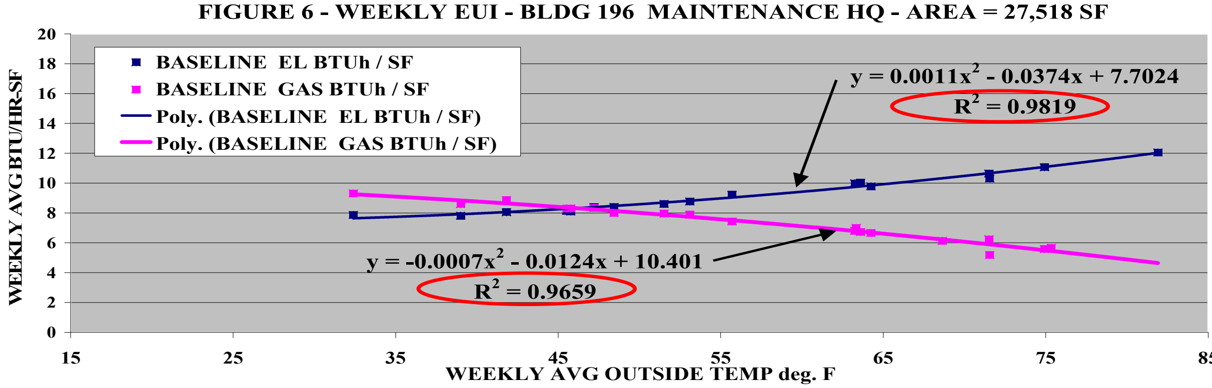
The remodel completed mid-January 2010. Post-remodel data showed a substantial improvement from baseline, until mid-November 2010. At that point, during an evening special event, indoor conditions got too warm. The work still being under warranty, the contractor was instructed to fix it. Those results (and club staff actions) being uncertain, the local HVAC shop became involved. Meanwhile, gas use had spiked above pre-remodel levels, see Figure 5.

Review of trend data from the BAS didn't show what had happened. Rumor was, the kitchen ventilation system had been used temporarily to supplement the main HVAC. Since the kitchen make-up and exhaust equipment wasn't monitored by the BAS, we asked, and were told that wasn't happening. After no progress for many weeks, we handed it over to the HVAC shop as a challenge, with extra vacation for whoever solved the problem. Shortly afterward, the HVAC shop discovered multiple problems with the gas-fired make-up air system and its controls. The problem was solved, and subsequent data showed some of the best efficiency we had seen since remodel completion. Two circles in Figure 5 show the last week of "failed" operation and the first week of corrected operation.

Example 3: Commissioning O&M Changes to Reduce Excessive Reheat

Early on, Building 196 showed high energy use, compared to peers within the original 10 buildings metered in 2009. Built in 2000, well insulated, with 83% office and 17% warehouse use, it should have been one of the most efficient of the 10, not a high outlier. Baseline gas and electric use, based on 20 weeks of data, are shown in figure 6.

FIGURE 6 - WEEKLY EUI - BLDG 196 MAINTENANCE HQ - AREA = 27,518 SF

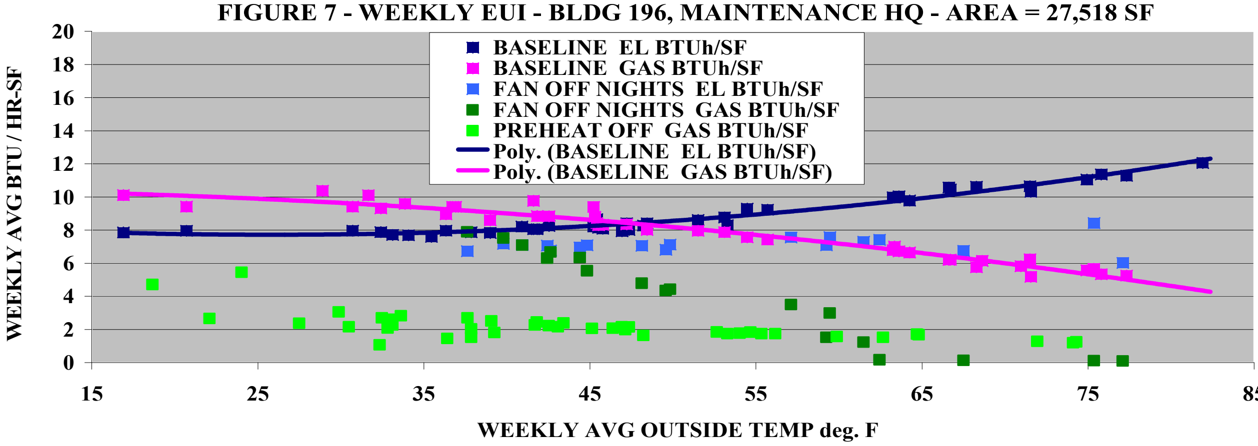


First, high summer gas use was conspicuous. Heat from lights and equipment should have sufficed above 60 deg. F. Also, note the high R^2 values for weekly "baseline" electricity and gas use fitted versus T_{out}. The advantage of using dis-aggregated heating and cooling is evident; ***total*** energy use would be almost flat, therefore showing little correlation with T_{out}.

When heating boilers were down for summer maintenance (T_{out} >80 F), gas use dropped below 0.5 BTUh/ft^2 (not shown), with cold complaints in many zones. Maybe the original design called for de-humidification. For whatever reason, the main air handler was first cooling to 51 F, then preheating to about 58 F. Also, the main air handler ran 24/7, although most of the building was a 6 AM to 4 PM occupancy. One small area was occupied until 10 PM, and a still smaller internal-gain-intensive interior zone, with stand-alone cooling, was occupied 24/7.

For O&M tune-up, we first tried turning the main HVAC fans off at night. This worked well for four months, until (apparently) failure to properly re-tension newly installed replacement fan belts caused belt failure. The HVAC maintenance supervisor ordered a return to 24/7 fan operation. Later, we persuaded the HVAC shop to mostly eliminate preheat. Results are shown in Figure 7, along with electrical use with fans off at night. Baselines are shown for comparison.

FIGURE 7 - WEEKLY EUI - BLDG 196, MAINTENANCE HQ - AREA = 27,518 SF



Both strategies greatly reduced gas use, but the reduced preheat was more effective. Reduced cooling load (from fan shutdowns and resulting partial cessation of preheat) also produced good electrical savings, as did reduction of preheat (not shown). Later, we installed air handler soft starts, so nightly fan shutdowns could resume. Diagnostic benchmarks had helped us to easily (1) spot a high energy user, (2) find a reheat problem, (3) test two ways to reduce reheat, (4) decide if both should be used, and (5) quantify an interactive effect, reduced cooling.

Example 4: Quick Response to a Problem

FIGURE 8 - WEEKLY EUI - Bldg 261 RAPCON, A = 17,800 SF

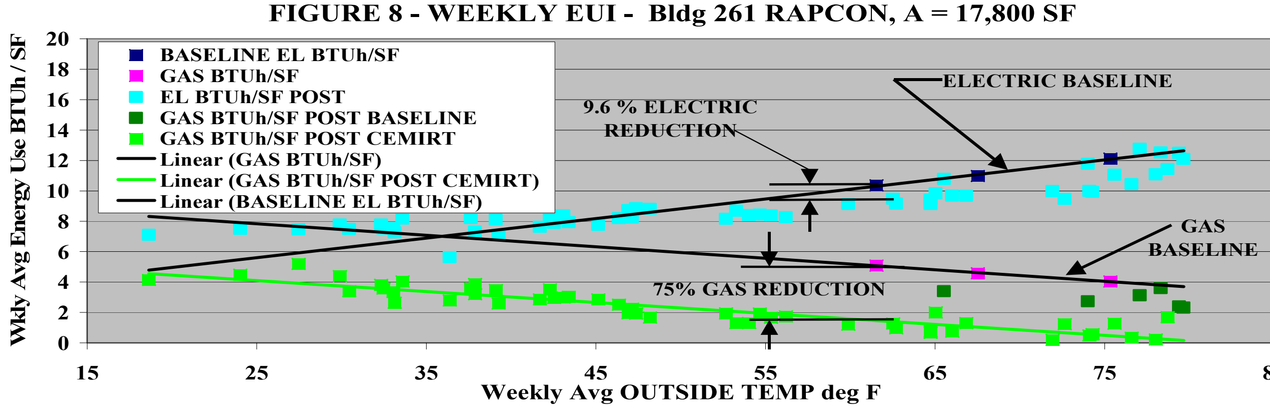


Figure 8 shows data for an air traffic control facility. This building had been completed and occupied in 2008, but not satisfactorily commissioned. The problems, going into a third summer, caused the HVAC shop to request an out-of-town repair specialist to fix a zone that couldn't be kept cool enough, causing critical equipment to be unreliable. With a few weeks before the specialist arrived, we learned the building had meters, and got to work.

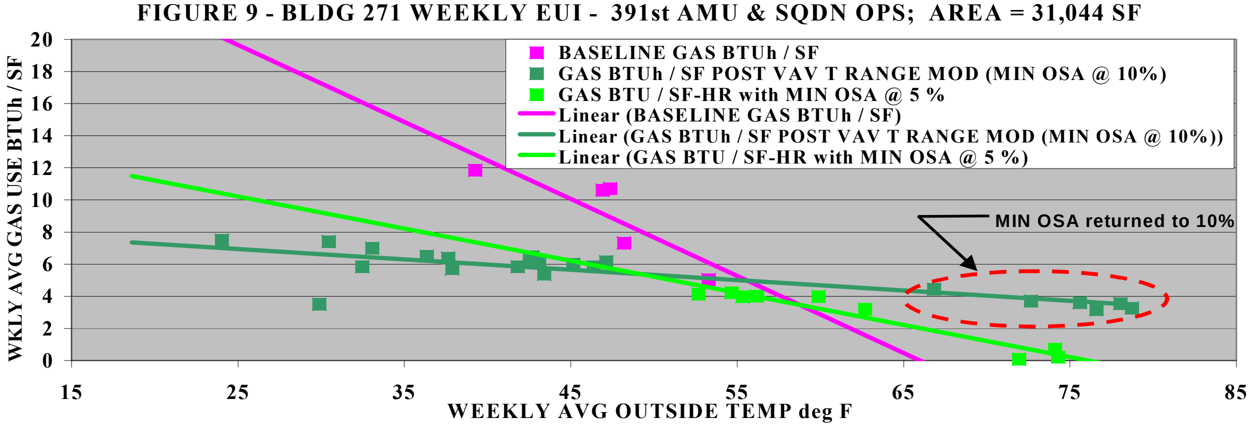
While collecting several weeks of baseline data, we also checked VAV box throttling ranges, other space temperatures, and CO_2. Space temperatures showed many zones were over-cooled. CO_2 showed gross over-ventilation that "diverted" half the chiller capacity. We recommended wider VAV throttling ranges and reduced minimum outside air (OSA). Upon arrival, the specialist ("Civil Engineering Measurement, Instrumentation and Repair Technician", or CEMIRT) seconded these suggestions, and found a broken OSA damper. After adjustments and repairs, the result was reduced OSA & reheat, 10% electrical and 75% summer gas savings, plus much improved equipment cooling, verified before the specialist left.

Example 5: Commissioning VAV Throttling Range and Outside Air

Building 271, completed about 2007, was another recently completed building that had not been thoroughly commissioned. The HVAC shop had gotten enough complaints that the building was considered "sensitive". Like other recent buildings at the base, it was well insulated and had a VAV system with reheat. The specified electric meter had been omitted from construction, but there was a gas meter. The energy office decided to see what could be done with just one part of the picture – but one we'd seen before.

Shortly after we started collecting baseline data, the EMCS shop discovered that someone had set maximum airflows for multiple zones such that there was no way the main air handler could deliver enough air. They responded (correctly) by cutting back VAV maximum airflows in some of the larger zones. See Figure 9. The energy office checked CO_2 and temperatures in some of the affected zones, as well as zones where occupancy was reportedly highly variable.

FIGURE 9 - BLDG 271 WEEKLY EUI - 391st AMU & SQDN OPS; AREA = 31,044 SF



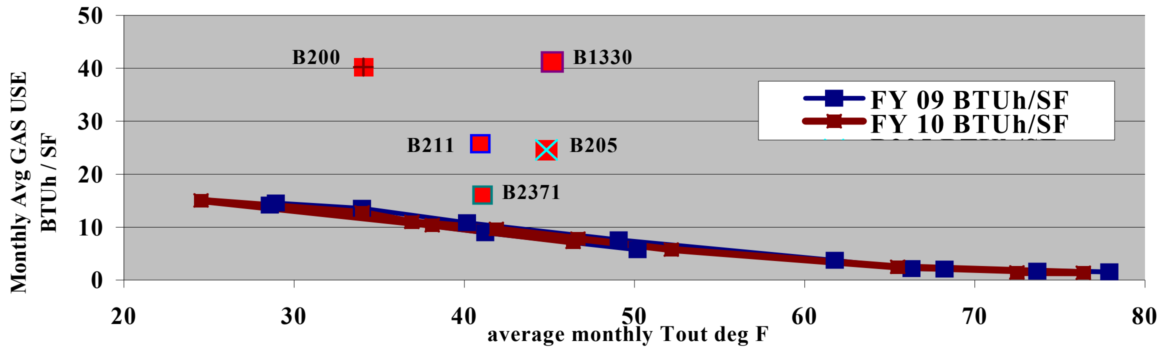
We found over-cooling and low CO2 in some areas, and recommended that VAV throttling ranges be increased by lowering the minimum VAV flows. However, the supply air temperature was still going over setpoint at times, due to high return air temperatures. We tried setting the minimum OSA damper at 5 % (from the prior value of 10%). CO_2 checks showed air quality was still good and return air temperatures indicated that the system was still getting lots of OSA somehow. We recommended an infrared thermography scan inside the plenum return in the attic during cold conditions to determine location of the OSA leaks.

Still later, the HVAC technician in charge of the building, preparing for an extended leave, belatedly realized the minimum OSA was at 5%. He raised it to 10% and left his understudy with an injunction not to change it back, then left without talking to anyone else. So we got 6 weeks of impaired savings. Based on what we had seen in other buildings, this probably increased the cooling load on a system that was at times "maxed out".

Example 6: Prioritizing Retro-Commissioning Effort by Peer Comparisons

New gas meters were being installed on over 28 additional large buildings. Considering where to start with the new buildings, we decided to compare the first weeks' data from new meters to a base-wide gas use profile we had created earlier. The results are shown in Figure 10.

FIGURE 10 - MHAFB BASE-WIDE GAS USE PATTERNS vs. "GAS HOGS"



The plot shows just the "high outliers" among newly metered buildings. This information immediately helped to establish our priorities - to focus on diagnosing high gas use at these buildings. Two buildings, 200 and 1330, proved to be process-dominated energy users, with process work-flow as an important independent variable, in addition to outside temperature. We had expected Building 1330, an aircraft-sized paint booth, to be a large gas user due to tempering of ventilation air. Building 200 also turned out to be ventilation-intensive, due to a need to limit indoor fuel vapor build-ups. Immediate effects of metering included beneficial "Hawthorne Effect" at both buildings, with estimated 20% savings. The top sergeants in both facilities started to more carefully limit ventilation runtimes, and schedule work for warmer periods during the day when possible, once they became more aware of equipment energy operating costs and associated management scrutiny.

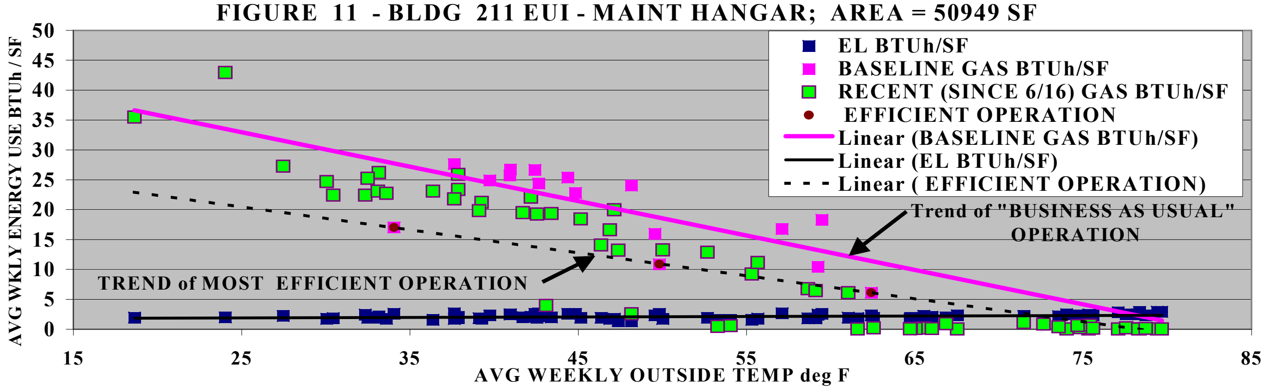
Building 200 was scheduled for ventilation system upgrades, consisting of addition of adaptive ventilation control (to limit kerosene fume concentration) using variable speed fans, and changes to damper controls that allowed partial recirculation (rather than 100% make-up air) when appropriate. We also recommended sharply curtailed use of under-floor radiant heat, which had no insulation below. Upgrade possibilities at Building 1330's paint booth were restricted, due to funding regulations, for which a waiver was required.

Buildings 205 and 211 are aircraft general maintenance hangars, dating to 1943 and about 1950, respectively. Both buildings are considered historic; after consulting base historic preservation personnel, we determined that roof insulation could be added at the next re-roofing, without unacceptably altering their appearance. Building 205 is also a "proxy" for 3 other nearly identical hangars, which are now in the queue for roof insulation as well. Both buildings also had operational problems with use of aircraft-sized hangar doors, see example 7.

Example 7: Identifying Inconsistent Operation

Hangar 211, a large-aircraft maintenance hanger, has aircraft doors at both gable ends of a bowstring truss uninsulated roof structure. Nineteen months of data are shown in figure 11.

FIGURE 11 - BLDG 211 EUI - MAINT HANGAR; AREA = 50949 SF



This data shows substantial scatter, indicating inconsistent operation. Baseline gas data shows overhead radiant heat "on" and the large doors intermittently open, year-round. The building manager had been warned not to let the hangar's wet-pipe fire sprinkler system freeze, and to let the HVAC department operate the heating. When aircraft were moved, or it got too hot, the doors were opened and heat remained on. There was no BAS in the building, and the HVAC department would not accept responsibility for operating the radiant heat's local controls.

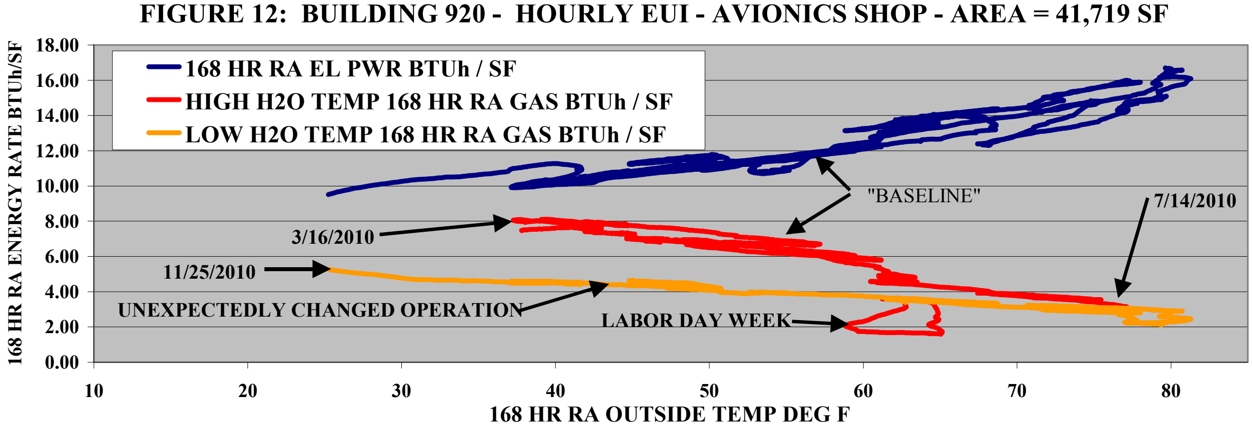
We got the building manager's orders changed so heat could be turned off during brief periods with doors open, and in the summer; this is represented by the points plotted in green. Long-term, the building is to be equipped with heating controls interlocked so that open doors preclude heating until near-freezing temperatures occur near the sprinkler piping.

To estimate the savings from roof insulation, we first assumed that the pre-insulation case for the building was represented by the trend-line labeled "most efficient operation" (which represents three weeks of operation when the doors were left closed during cold weather). Then we compared that trend to the trend of a similar hangar which had had roof insulation installed.

Example 8: Example Data with Hourly Time Steps – and Dueling Thermostats

With hourly data, a continuous benchmark is possible, with energy use and T_{out} averages advancing simultaneously in one-hour time steps. Figure 12 shows an example.

FIGURE 12: BUILDING 920 - HOURLY EUI - AVIONICS SHOP - AREA = 41,719 SF



Building 920, used to service aircraft electronics, was a high energy user. Hourly data showed both substantial electrical use and summer gas use. Irregular electrical use is from changing occupancy. Built about 2000, the building was well insulated. With its electrical use, it should have needed little heat. The chiller could be heard running while walking on ice to read the meter. Some critical special equipment was never turned off, and needed ambient temperatures of 70 F +/- 5 degrees, 24/7, for stable operation, so setback wasn't an option.

Following some seasonal maintenance, we saw gas use trending below baseline, with no explanation. We assumed the HVAC shop had made a previously requested change. After we thanked them for finally making the requested change, they responded, "What change?"

The slope of the baseline gas use curve (or the new trend) can be interpreted as building loss coefficient divided by heating system efficiency. So the data in figure 12 implies a substantial improvement in heating system efficiency (or a substantial reduction in heating system losses). Whatever was happening was worth investigation, and possibly replication.

A review of BAS trends and settings showed the only change associated with the large drop in gas use was reduced summer-time boiler water temperature, by 40 degrees F. True, lower boiler temperature reduces stack losses, but should only have produced a few percent of boiler efficiency improvement, not up to a 40 % reduction in gas use. What else was happening?

Further trend reviews showed that several zones were alternately heating and cooling (definitely a loss mechanism), but the frequency of alternation had slowed substantially with lower heating water temperatures; lowered heating water temperature produced less frequent overshoots, and reduced losses. Further review showed that many zones had the same heating and cooling setpoints (70 F) with zero dead-band! We also found a single zone HVAC unit serving two large work bays, separated by a hallway, with one thermostat, and lots of reheat.

Conclusions and Lessons Learned

Diagnostic benchmarks proved indispensable for all phases of our commissioning work. The examples show these benchmarks can quickly detect trends and trend changes in energy use. Using cyclic averages of energy use, such trending has the fastest possible response time.

Other than rigorous sampling, diagnostic benchmarks presented here differ from previously described work in three ways. They rely only on outside temperature as the independent variable. Deliberately keeping heating and cooling energy uses as separate dependent variables improves correlation with outside temperature. Separate consideration of demonstrably atypical data – statistical heresy – further improves robustness of trends.

Technological capability is limited by the measurements that support it. This substantial improvement in energy measurement capability enables many improvements in building commissioning technology. Many other uses are also possible, as a result of capabilities shown.

Commissioning in New or Existing Buildings

Commissioning new buildings. With performance targets from simulations, diagnostic benchmarks show if targets are met. Weather normalization, and re-running simulations for changed occupancies/schedules, makes these two issues moot. Remaining reasons for shortfalls are lack of skill of the model or modeler, or defects in building design, construction or O&M that need correction. People writing contracts for construction need to change contracts to accommodate diagnostic benchmarking. Thus, most of the first three goals in the introduction are satisfied.

For commissioning new buildings with no simulated targets, we'd use peer comparisons, i.e. "perform as well as that building" or "use much less energy than base-wide average".

Commissioning existing buildings. Goal setting depends on the type of change being made. For capital upgrades, the goal should be decreased energy use at least matching what justified the expenditure. Success is shown by a sufficiently lowered benchmark. For O&M changes, goal setting can be as simple as "lower is better, lowest is best". Precisely pre-calculating expected savings is an unneeded hurdle. Make plausible changes, observe changed benchmarks, and learn; then decide which measures to make permanent. Also, during continuous commissioning, optimizing O&M makes plant shortcomings more obvious, and helps with planning and right-sizing future retrofit equipment. This feedback - what works - finishes meetings our third goal.

Savings retention. The last two goals set out were to keep hard-won savings, and convince financiers and regulators that efficiency warrants their continued support. Space didn't permit showing numerous examples where equipment or operational problems occurred, without loss of comfort. Continued benchmarking constrains losses when equipment or people fail.

Finally, diagnostic benchmarks provide an opportunity for the community of energy efficiency professionals – researchers, designers, modelers, builders, evaluators, and those who operate and maintain buildings – to more conclusively demonstrate their professionalism. The building community, by showing visibly improving efficiency trends, can persuade those who write checks and regulations to do so more confidently - and we will all prosper thereby.

Other Uses

- **Modeling.** Making careful comparisons of modeled to measured outcomes could show us how much reliance on models is warranted. It's feedback we need.
- **Building Codes.** Building codes should require metering to support diagnostic benchmarking in new buildings, and its use in new and existing buildings. For example, San Francisco presently requires buildings over 10,000 square feet to use ESPM.
- Building codes could set annual EUI targets, based on diagnostic benchmarks proved achievable. Also, proposed "stretch codes", sometimes used to test future code upgrades, could be field-evaluated using diagnostic benchmarking prior to adoption.
- **ISO 50001 and IPMVP.** These standards presently do not include use of diagnostic benchmarking. Diagnostic benchmarking should be considered for addition to both.
- **Evaluation.** Process and impact evaluations mandated by regulatory requirements could be more precise, less expensive, require smaller samples, and provide faster results, by use of diagnostic benchmarks both before and after treatments.
- **ESPM.** Usually voluntary, this benchmarking method is mandatory for most US Federal facilities. Many federal facilities, such as military ones, are mandated to use automated metering ("AMR") in larger buildings. EPA should consider upgrading the weather normalization part of ESPM to use hourly data, as a convenience to users with AMR.
- **Automated Fault Detection and Diagnosis (FDD)**. Manufacturers of building automation systems (BAS) could upgrade their product offerings from BAS to true Energy Management and Control Systems (EMCS), by using diagnostic benchmarks.
- **LEED.** It's recommended that LEED adopt diagnostic benchmarking as a quality assurance method, both for new and for existing buildings.
- **Building energy research.** Some building science research questions depend on modeling, because testing has been too expensive or too imprecise. Some of these issues can now be field tested. For example: Oregon code allows "conditioned crawlspaces" (insulated only at stem-walls) for new homes. Post-occupancy, they become eligible for a subsidized under-floor insulation retrofit. Does this make sense? Similarly, perimeter-

insulated residential slabs on grade are allowed in most places, but in Alaska, heat losses are so large that permafrost can melt, in which case insulation is required under the whole slab. At what outdoor temperature(s) is full under-slab insulation appropriate? How do you quickly measure benefits of a reflective roof retrofit?

Lessons Learned

Diagnostic benchmarks probably have more uses that have been overlooked. If you wish to try diagnostic benchmarks on your next project, here are some additional pointers:

- For new buildings, make sure that the metering and weather data collection are well in hand, early. Commission the meters and the data collection system early.
- Diagnostic benchmarks serve contractual needs with a single fuel. But their diagnostic power improves when heating and other energy uses are recorded on different meters.
- Make sure that your contract will facilitate using diagnostic benchmarks, provide the right motivations, and not tie everyone's hands.
- For existing buildings, do O&M first, then consider capital upgrades.
- For O&M, you must have control over changes in the building sequence of operation.
- If you are a researcher or innovator, diagnostic benchmarking enables some very challenging measurements; here's a possible way to prove your favorite idea works.

Acknowledgements

The author wishes to thank the US Air Force and Tetra Tech EMI for allowing the author freedom to employ and demonstrate the methods described here. In particular, thanks to the 366th Civil Engineering Squadron at Mountain Home AFB, Idaho, the Base Energy Manager, Joseph Armstrong, and the base EMCS shop, led by Thomas Howe. Special thanks are also due to the HVAC shop, for their willingness to consider and even try what they saw as risky new ideas.

References

[EPA] U.S. Environmental Protection Agency. 2011. *ENERGY STAR Portfolio Manager Methodology of Accounting for Weather* http://www.energystar.gov/ia/business/evaluate_performance/Methodology_Weather_20110224.pdf. Washington, DC.: U.S Environmental Protection Agency

Goldberg, Miriam L. 1980. *Report No. 105 - Residential Energy Consumption: Use of Aggregate Data to Measure Physical Properties of Housing Stock*. May 1980. Princeton, NJ.: Princeton University Center for Energy & Environmental Studies.

Lambert, Les. 1998. – "Energy Signature Analysis: Radar for Energy managers." In *Proceedings of the ACEEE 1998 Summer Study on Energy Efficiency in Buildings*, 3:217-228. Washington, DC.: American Council for an Energy-Efficient Economy.